INFORMATION AND CONTROL 18, 257–260 (1971)

A Note on the Summability of the Entropy Series*

J. KEILSON

Department of Statistics, The University of Rochester, Rochester, New York 14627

INTRODUCTION

The *entropy* of a denumerable probability distribution $p = \{p_n; n = 1, 2,...\}$ is defined as

$$H_p = \sum_1^\infty p_n(-\log p_n), \tag{1}$$

where $p_n \log p_n$ is defined as 0 whenever $p_n = 0$ [see, for example, Feinstein (1958)]. The function H_p is a sum of nonnegative terms and consequently invariant under rearrangements of the terms. The probabilities p_n may therefore be reordered so that they are monotonic decreasing (nonincreasing), to facilitate discussion of the finiteness of H_p. All terms may also be assumed positive, for nontriviality. We hereafter confine attention to monotonic nonincreasing sequences of positive probabilities $p = \{p_n\}$. (Actually, ultimate monotonicity is enough.)

In this note we present two sufficient conditions to assure the finiteness of the entropy H_p for such distributions p. They are given in the two theorems below. The proofs are based on corollaries to Lemmas 1 and 2, respectively; these lemmas and corollaries may have some independent interest.

That H_p need not be finite may be seen from the example

$$p_n = \frac{K}{n \log^2 n}$$

where K is a normalization constant. It is readily seen that p_n is summable but that $-p_n \log p_n$ is not.

The two theorems are:

THEOREM 1. *If, for some* $\epsilon > 0$, $E(N^\epsilon) = \sum p_n n^\epsilon < \infty$, *then* $H_p < \infty$.

* Supported, in part, by the Office of Naval Research under Contract N000 14-68-A 0091.

THEOREM 2. *If $q_n = np_n$ is a slowly varying sequence in the sense that $q_{kn}/q_n \to 1$, as $n \to \infty$, for all positive integers k, and if $E[\log N] = \sum p_n \log n < \infty$, then $H_p < \infty$.*

The first states that the finiteness of any fractional moment of p is sufficient, while the second states that for sufficiently regular sequences, the finiteness of $E(\log N)$ is sufficient. An example for Theorem 2 is provided by $p_n = K(n \log^3 n)^{-1}$ for which H_p is finite.

LEMMA 1. *Let r_n, $n = 1, 2, 3, \ldots$, be a monotone decreasing sequence for which $r_n \to 0$, $n \to \infty$. Let $f_n \geqslant 0$ be such that $\sum_1^\infty f_n r_n < \infty$. Then $r_n \sum_1^n f_j \to 0$, as $n \to \infty$. (r_n need not be summable.)*

Proof. Represent $r_n = \sum_n^\infty \zeta_j$ where $\zeta_j = r_j - r_{j+1} \geqslant 0$. Then

$$\sum_1^\infty f_n r_n = \sum_1^\infty f_n \sum_n^\infty \zeta_k = \sum_1^\infty \zeta_k \sum_1^k f_j < \infty,$$

and

$$\sum_n^\infty \zeta_k \sum_1^k f_j \to 0, \qquad n \to \infty. \tag{2}$$

Moreover,

$$\left\{\sum_1^n f_j\right\} r_n = \left\{\sum_1^n f_j\right\} \sum_n^\infty \zeta_k \leqslant \sum_n^\infty \zeta_k \sum_1^k f_j$$

and the lemma then follows from (2). Setting $f_n \equiv 1$ and $r_n = p_n$, we have

COROLLARY 1. *$np_n \to 0$ as $n \to \infty$.*

Suppose now that $\sum n^\epsilon p_n < \infty$ for some $\epsilon > 0$. The lemma then implies that $p_n \sum_1^n j^\epsilon \to 0$. Since $\sum_1^n j^\epsilon \sim n^{1+\epsilon}/(1+\epsilon)$, we have

COROLLARY 2. *If $E(N^\epsilon) = \sum n^\epsilon p_n < \infty$ for some $\epsilon > 0$, then $n^{1+\epsilon} p_n \to 0$.*

Remark. Note that only the behavior of the sequence $\{r_n\}$ or $\{p_n\}$ for sufficiently large n is of importance and we need only require that the sequence be monotone ultimately.

Proof of Theorem 1. It is well known that $-x \log x \to 0$, $x \to 0+$. Consequently for any $\delta > 0$, $-x^\delta \log x^\delta \to 0$, and hence $-x^\delta \log x \to 0$, as

$x \to 0+$. It follows that $-\log x < c_\delta x^{-\delta}$, $0 < x \leqslant 1$, for some $c_\delta > 0$, and

$$H_p \leqslant c_\delta \sum_1^\infty p_n^{1-\delta}.$$

From Corollary 2, moreover, we will have $p_n < c_\epsilon n^{-(\epsilon+1)}$, for some $c_\epsilon > 0$. Hence,

$$H_p \leqslant c_\delta c_\epsilon^{1-\delta} \sum_1^\infty \frac{1}{n^{(1+\epsilon)(1-\delta)}}.$$

We may choose δ small enough so that $(1 + \epsilon)(1 - \delta) > 1$ and the theorem follows.

LEMMA 2. *Let $L(x)$ be a slowly varying function on $(0, \infty)$ and let $U(x) = x^\zeta L(x)$, $0 < \zeta < \infty$. If $U(x)$ is monotonic increasing with x, then for any $\epsilon > 0$,*

$$H(x) = \frac{U(x)}{x^{\zeta+\epsilon}} \to 0, \; x \to \infty.$$

For the definition of slowly varying functions, see Feller (1966).

Proof. From our assumptions,

$$2^{-k\epsilon} L(2^k) = L(1) \prod_{j=1}^{k} \{\theta_j 2^{-\epsilon}\},$$

where $\theta_j = L(2^j)/L(2^{j-1}) \to 1$, as $j \to \infty$. Hence $2^{-k\epsilon}L(2^k) \to 0$ as $k \to \infty$. Moreover, for $2^{M-1} \leqslant x \leqslant 2^M$, we have

$$U(2^{M-1}) \leqslant U(x) \leqslant U(2^M)$$

and

$$\frac{1}{2^{M(\zeta+\epsilon)}} \leqslant \frac{1}{x^{\zeta+\epsilon}} \leqslant \frac{1}{2^{(M-1)(\zeta+\epsilon)}}.$$

Hence

$$\frac{1}{2^{\zeta+\epsilon}} H(2^{M-1}) \leqslant H(x) \leqslant 2^{\zeta+\epsilon} H(2^M),$$

and $H(x) \to 0$, as $x \to \infty$.

COROLLARY. *If p_n is a positive monotone decreasing sequence of probabilities, and np_n is slowly varying, then for any $\epsilon > 0$*

$$\frac{1}{n^{1+\epsilon}p_n} \to 0, \quad n \to \infty.$$

For, if np_n is slowly varying, so also is $1/(np_n)$ and $1/p_n$ is monotonically increasing. The corollary then follows from the lemma. We may now prove Theorem 2.

Proof of Theorem 2. From the corollary above

$$\frac{1}{n^{1+\epsilon}p_n} < K,$$

for some $K > 1$. Hence $-\log p_n < \log K + (1 + \epsilon) \log n$ and

$$\sum_1^\infty p_n(-\log p_n) < (1 + \epsilon) \sum_1^\infty p_n \log n + \log K.$$

The theorem then follows.

ACKNOWLEDGMENT

I would like to thank W. J. Hall, E. van der Meulen, and G. S. Mudholkar for their interest and suggestions.

RECEIVED: July 15, 1970

REFERENCES

FELLER, W., "An Introduction to Probability Theory and its Applications, Vol. II" Wiley, New York, 1966.

FEINSTEIN, A., "Foundations of Information Theory," McGraw-Hill, New York, 1958.